

Sémantický web – 10 let poté

Vilém Sklenák – sklenak@vse.cz

Vysoká škola ekonomická, fakulta informatiky a statistiky,
katedra informačního a znalostního inženýrství

Inforum2011, 26. 5. 2011

Obsah

- 1 Úvod
- 2 Jak to začalo
- 3 Jak to pokračovalo
- 4 Jak to je nyní
- 5 Co dál
- 6 Závěr

Jak začala historie

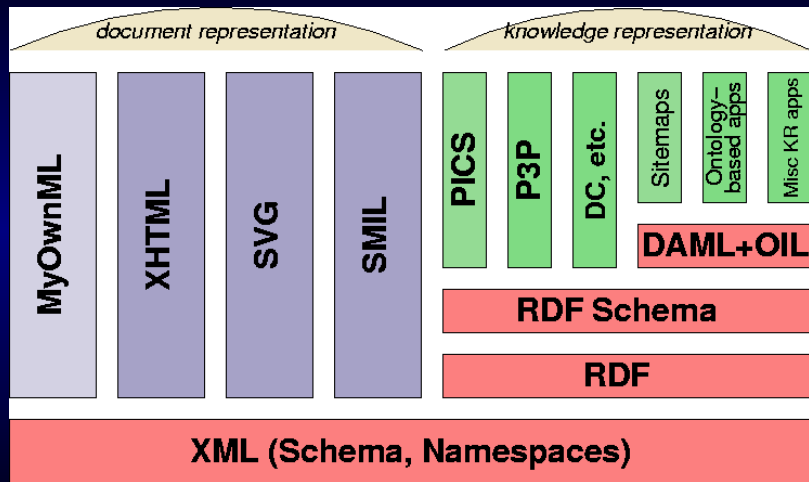
- Vznik myšlenky a rozvoj základních principů sémantického webu není věcí posledních 10 let.
- Ale – 2001 = Berners-Lee, Tim, Hendler, James, Lassilla, Ora. The Semantic Web. *Scientific American*, 2001. vol. 284, no. May. str. 35–43. URL: <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.
- V té době však práce na sémantickém webu trvaly již několik let.
- Od počátku platí, že *„sémantický web je rozšířením současného webu, jež datům přiřazuje přesný význam, díky kterému bude možná kooperace jak lidí, tak softwaru“*.

Úmysly

- Sémantický web není nějaký nový web, ale jde o rozšíření konceptu a doplnění dat toho stávajícího.
- Doplnění o metadata, která by měla popisovat sémantické informace webových zdrojů a která by měla být zápsána pomocí strojově srozumitelných jazyků.
- Součástí metadat by také byla použitá slovní zásoba a soubor vztahů mezi jednotlivými pojmy.

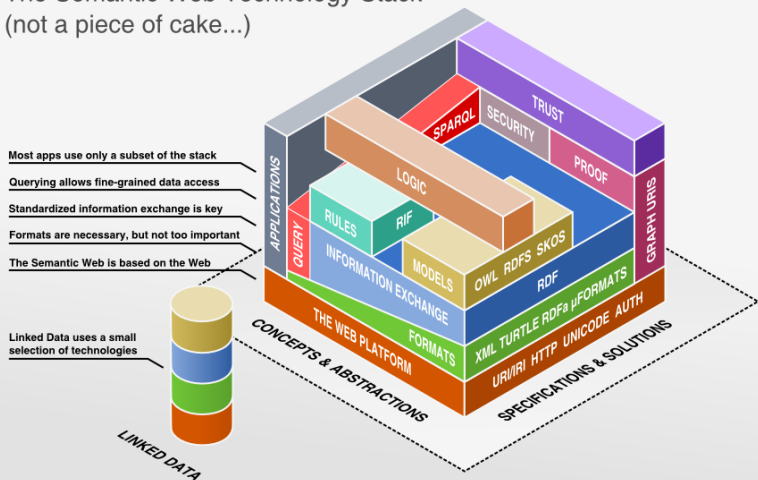


Jazyky sémantického webu



Technologie a jazyky sémantického webu

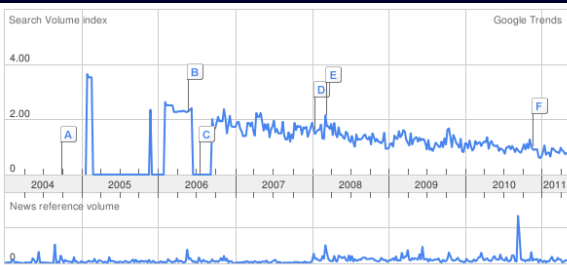
The Semantic Web Technology Stack
(not a piece of cake...)



Rok po roku

- 2000 – <http://semanticweb.org>
- 2001 – <http://www.w3.org/2001/sw/>
- 2002 – poprvé *International Semantic Web Conference*
- 2004 – standardy *RDF, OWL, RDFS*
- 2007 – standardy *GRDDL (Gleaning Resource Descriptions from Dialects of Languages), SAWSDL (Semantic Annotations for WSDL and XML Schema)*
- 2008 – standardy *SPARQL Query Language for RDF, RDFa*
- 2009 – standardy *SKOS (Simple Knowledge Organization System), POWDER (Protocol for Web Description Resources)*
- 2010 – standardy *RIF (Rule Interchange Format)*

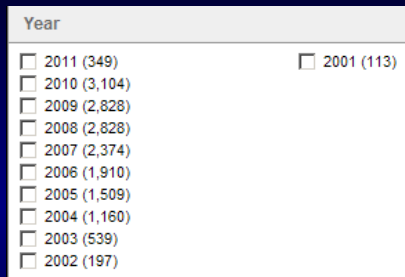
Google a dotaz „semantic web“ v čase



- A** [Berners-Lee says cooperation needed for "Semantic Web"](#)
USA Today - Sep 30 2004
- B** [Semantic Web will drive Internet](#)
Macworld UK - May 25 2006
- C** [Google: Semantic Web must overcome incompetence](#)
Builder UK - Jul 19 2006
- D** [Semantic Web takes big step forward](#)
Computerworld - Jan 16 2008
- E** [Yahoo Search takes aim at Semantic Web](#)
Computerworld - Mar 13 2008
- F** [Elsevier Sponsors 2010 Semantic Web Challenge](#)
MarketWatch - Nov 16 2010

Sémantický web v publikacích

Field: Publication Year	Record Count	% of 10584	Bar Chart
1999	4	0.0378 %	
2000	27	0.2551 %	
2001	125	1.1810 %	
2002	297	2.8061 %	
2003	526	4.9698 %	
2004	842	7.9554 %	
2005	1150	10.8655 %	
2006	1326	12.5283 %	
2007	1548	14.6259 %	
2008	1805	17.0540 %	
2009	1813	17.1296 %	
2010	944	8.9191 %	
2011	177	1.6723 %	



2001	37	2006	227
2002	74	2007	244
2003	109	2008	284
2004	142	2009	256
2005	186	2010	409

Prestižní konference

- *International Semantic Web Conference (ISWC)* – <http://iswc2011.semanticweb.org>
- a *Extended Semantic Web Conference (ESWC)* – <http://www.eswc2011.org/>.

Struktura konferencí:

- 1 výzkum (research papers),
- 2 použití (semantic web in use),
- 3 a využití v komerční sféře (industry track).

ISWC2011 – hlavní témata

- Management of Semantic Web Data
 - Languages, tools, and methodologies for representing and managing Semantic Web data
 - Database, IR, and AI technologies for the Semantic Web
- Natural Language Processing
 - Machine learning and information extraction for the Semantic Web
- Ontologies and Semantics
 - Specific ontologies and ontology patterns for the semantic web
- Semantic Web Engineering
 - Methods for Semantic Web application development
- Social Semantic Web
 - Social networks and processes on the Semantic Web
- User Interfaces to the Semantic Web
 - Interacting with Semantic Web data

Aplikace

- 1 Freebase – firma Metaweb, akvizice Googlu v roce 2010
- 2 GetGlue – služba typu vytváření sociálních sítí v oblasti zábavy.
- 3 FlipBoard – „sociálně“ orientovaný časopis, který má integrovány sémantické přístupy s cílem lepšího určování relevance informací.
- 4 Hunch – personalizované doporučování v oblasti volného času (filmy, knihy, dovolená aj.) s využitím technik mapování a rozhodovacích stromů.
- 5 Apture – vyhledávací služba založená na sémantice kontextu.

Semantic Enterprise: What Are The Gorillas Doing? (Oracle, IBM, HP, Cisco, Microsoft and SAP)

By Bernard Lunn on August 3, 2010 7:00 AM



In [Crossing The Chasm](#) terminology, “gorillas” are the dominant vendors. Simple message for start-ups - don't mess with them!

In this post, we want to understand what the gorillas are doing to apply semantic web technology to the enterprise. The gorillas in this market are: Oracle, IBM, HP, Cisco, Microsoft and SAP.

Firmy

Hepp Research GmbH

Experts in articulating your value proposition on the World Wide Web

[Profile](#)
[RDFa for Google](#)
[Products](#)
[Training](#)
[Consulting](#)
[Contact](#)
[Imprint](#)

Semantic SEO for Google with GoodRelations and RDFa

Breaking News: Google has just announced that they now support the [GoodRelations vocabulary](#) developed by Prof. Hepp for Google Rich Snippets.

If you add a few extra lines to the HTML code of your Web site, **you will feed Google and hundreds of evolving mobile and recommender applications** in one turn.

On this page, we provide **slightly modified markup** that is **superior to the [original Google pattern](#)** for marking up your pages. Our recipe, developed by the inventor of GoodRelations for you personally, is understood by ALL RDFa-aware search engines, shopping comparison sites, and mobile services, while the Google recipe is perfect for Google only.

Principle and Effect

If you add a little bit of additional mark-up to your page, Google will use that information to substantially enhance the rendering of your page directly in the search results list.

There is preliminary evidence that search results with respective extensions get **a 30 % higher click-through rate (CTR)**.

The flickering nature of search results in the new Google Instant technology will very likely further increase the gains of more appealing rendering in the results list. So don't put this off. **Act today!**

ORACLE ([Sign In/Register for Account](#) | [Help](#)) United States ▾ Communities ▾ I am a... ▾ I want to... ▾

Products and Services Downloads Store Support Training Partners About

Oracle Technology Network > Database > Database Options > Semantic Technologies

Database 11g
Database Focus Areas
Database Features
Database Options
Database Express Edition
Database Upgrade
Berkeley DB
TimesTen In-Memory Database
Exadata
Database 10g
Database Lite
RDB
Audit Vault
Secure Backup
Real Application Clusters
Clusterware
Cloud Storage
Multimedia

Oracle Database Semantic Technologies

Overview ▶ Downloads ▶ Documentation ▶ Community ▶ Learn More

Partners ▶ Demos ▶ Training ▶ Performance ▶ Standards & Open Source ▶ Conference Presentations ▶ Success Stories ▶ News

Oracle Database 11g Semantic Technologies is an open, standards-based, scalable, secure, reliable and performant RDF management platform. Based on a graph data model, RDF data (triples) are persisted, indexed and queried, like other object-relational data types.

Application developers use the power of the Oracle Database 11g to design and develop a wide range of semantic-enhanced business applications in areas that include intelligence, law enforcement, integrated bioinformatics and health care informatics, finance, web social network, and media, games, and content management.

[Learn How Semantic Technologies Makes Enterprise Data Smarter...](#)

Oracle Database Semantic Technologies Feature Overview (PDF 70 KB)

Linked Data

- jeden z aktuálních trendů směřujících k budování sémantického webu
- Tim Berners-Lee: *„The value of your own information is very much a function of what it links to, as well as the inherent value of the information within the web page“*
- propojování strukturovaných dat na webu pomocí odkazů
- styl publikování a propojování, soubor doporučení i data samotná
- *web dat*

W3C – projekt LOD

- od r. 2007
- vytipovat datové zdroje, publikované pod otevřenou licencí, a zpřístupnit je na webu za použití rámce RDF podle principů linked data
- nejprve malé skupiny, malá data, dnes však i BBC, Thomson Reuters, Kongresová knihovna aj.
- <http://data.gov.uk/>,
- <http://www.bl.uk/bibliographic/datafree.html>
- ...

Linked Data – principy

- 1 URI – globální unikátní identifikátory (v rámci jmenného prostoru celého webu), perzistentní,
- 2 HTTP – standardní protokol pro komunikaci pro WWW, podle hlavičky požadavku lze vracet formát podle kontextu,
- 3 RDF – nativní datový model objekt-predikát-subjekt, současně umožňuje přiřazovat význam odkazům,
- 4 ontologické jazyky RDFS, OWL, SKOS.

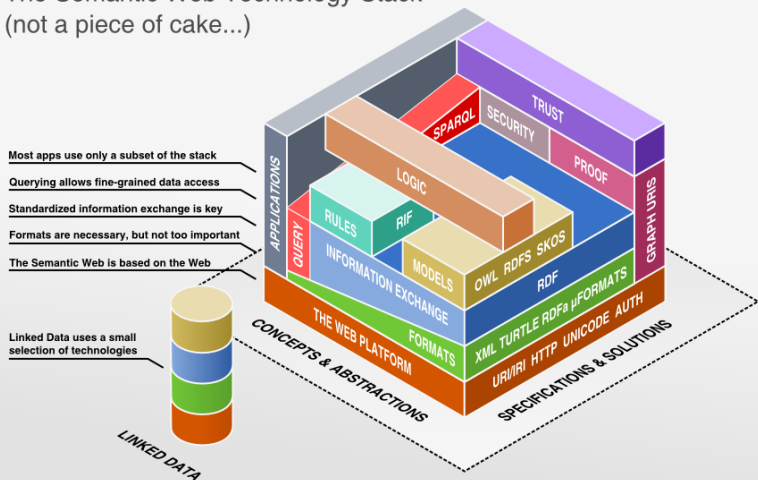
Linked Data a sémantický web

Linked data i sémantický web se shodují v mnoha předpokladech: snaží se o rozšíření současného webu, propagují použití rámce RDF, identifikátorů URI a efektivnější zpracování dat webovými aplikacemi. Také propojování dat je společným bodem obou iniciativ.

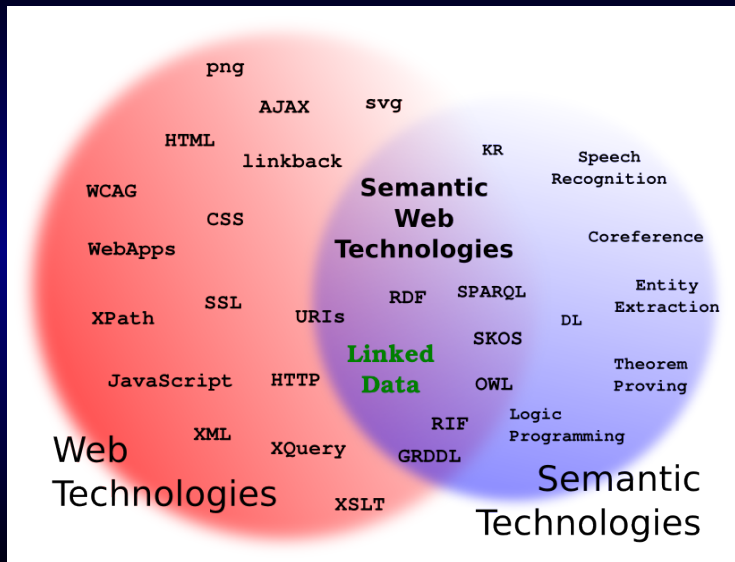
„Vize sémantického webu je založena na budování globálního webu strojově zpracovatelných dat. Zatímco sémantický web, nebo web dat je cílem či výsledkem, linked data představuje prostředek nebo způsob jeho dosažení.“

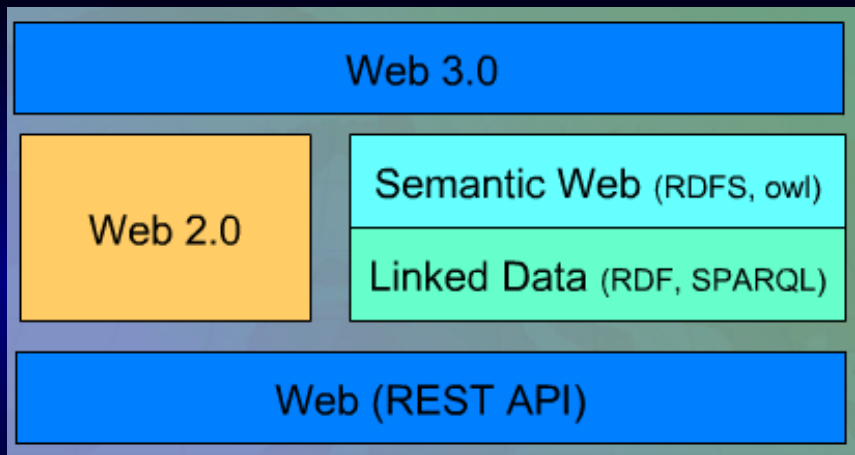
Linked Data a sémantický web

The Semantic Web Technology Stack
(not a piece of cake...)



Linked Data a sémantický web





Děkuji za pozornost.